16.002 Lecture (8)

Communication Systems
Amplitude Modulation
(AM)

April 9, 2008

Today's Topics

1. Amplitude modulation of signals
2. Application issues

Take Away
Fourier Transform methods facilitate the understanding of modulation in communication systems

Required Reading
O&W-8.0, 8.1, 8.2, 8.3, 8.7

Communication systems typically transmit information that has content at relatively low frequencies by encoding it, in one fashion or another, onto carrier signals at much higher frequencies. For example, amplitude modulated (AM) commercial radio broadcasting systems typically transmit voice and music signals using electromagnetic waves that pass readily through the atmosphere. The voice and music signals have frequency content typically in the range of about 20 Hz (cycles per second) up to about 20 kHz (thousands of cycles per second). The physical characteristics of the atmosphere make it very difficult for signals at these frequencies to be transmitted at distances beyond a few meters. Rather, these signals are encoded onto high frequency sinusoidal carrier signals that are in the range of 520 kHz up to 1.75 MHz (millions of cycles per second), by modulating the amplitude of the carrier wave. Typically the frequency of the carrier is one or more orders of magnitude greater than the frequency of the signal it is "carrying". Fourier transform methods are an ideal means for understanding the workings of these kinds of communication systems.

Amplitude Modulation

We will start our study of communications systems by first analyzing the AM method of communicating signals. The two most common forms of amplitude modulation use either a complex carrier or a single sinusoidal carrier. We will consider each of these in turn.

Complex Carrier Modulation

A complex carrier signal *c(t),* at a carrier frequency ω_c, is described mathematically as the complex exponential

$$c(t) = e^{j(\omega_c t + \theta)}$$

For convenience we choose the initial time so that the phase (θ) is zero. Then, if *x(t)* is the signal or information that is to be transmitted by the carrier, the signal *x(t)* is encoded onto the carrier by multiplying the carrier by *x(t)*

$$y(t) = x(t) \cdot c(t) = x(t)\, e^{j\omega_c t}$$

which is to say, the carrier's amplitude is modulated by the signal *x(t)*. Now we know that multiplication in the time domain is equivalent to convolution in the frequency domain. Thus, the Fourier transform of the signal *y(t)* is the convolution of the Fourier transforms of *x(t)* and *c(t)*

$$Y(j\omega) = X(j\omega) * C(j\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\theta)\, C(j(\omega-\theta))\, d\theta$$

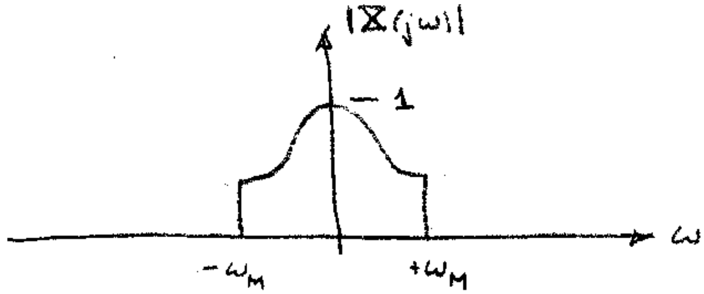
Earlier we took the Fourier transform of a complex exponential and determined it is a delta function

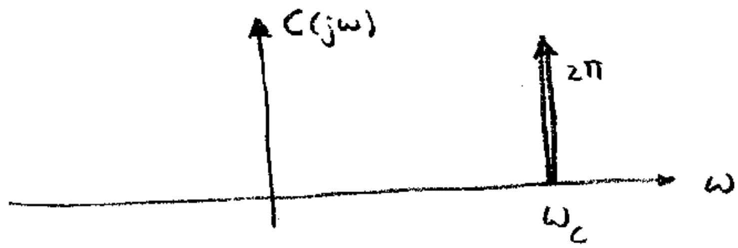
$$C(j\omega) = 2\pi\, \delta(\omega - \omega_c)$$

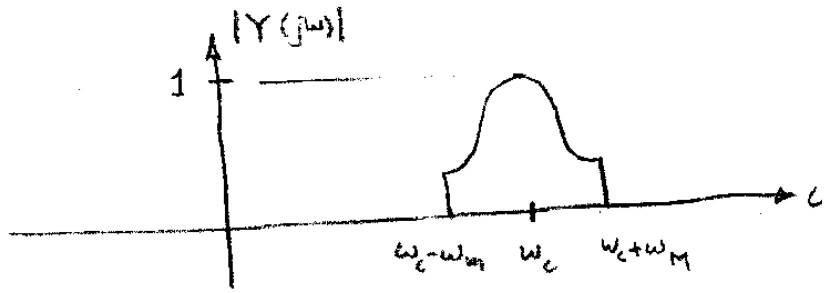
and upon substitution into the convolution equation we obtain

$$Y(j\omega) = X(j(\omega - \omega_c))$$

Thus, as a result of modulation, the transform of the signal *x(t)* is shifted on the frequency axis by the carrier frequency. We can visualize the situation by considering the magnitude of $X(j\omega)$. We suppose that the signal *x(t)* is a real function of time and that its frequency content is bounded by some maximum frequency ω_M. Hence, all of the signal power lies in the range $\pm\omega_M$, as depicted in the first figure below. The second figure depicts the delta function at ω_c and the third figure shows the result of amplitude modulation





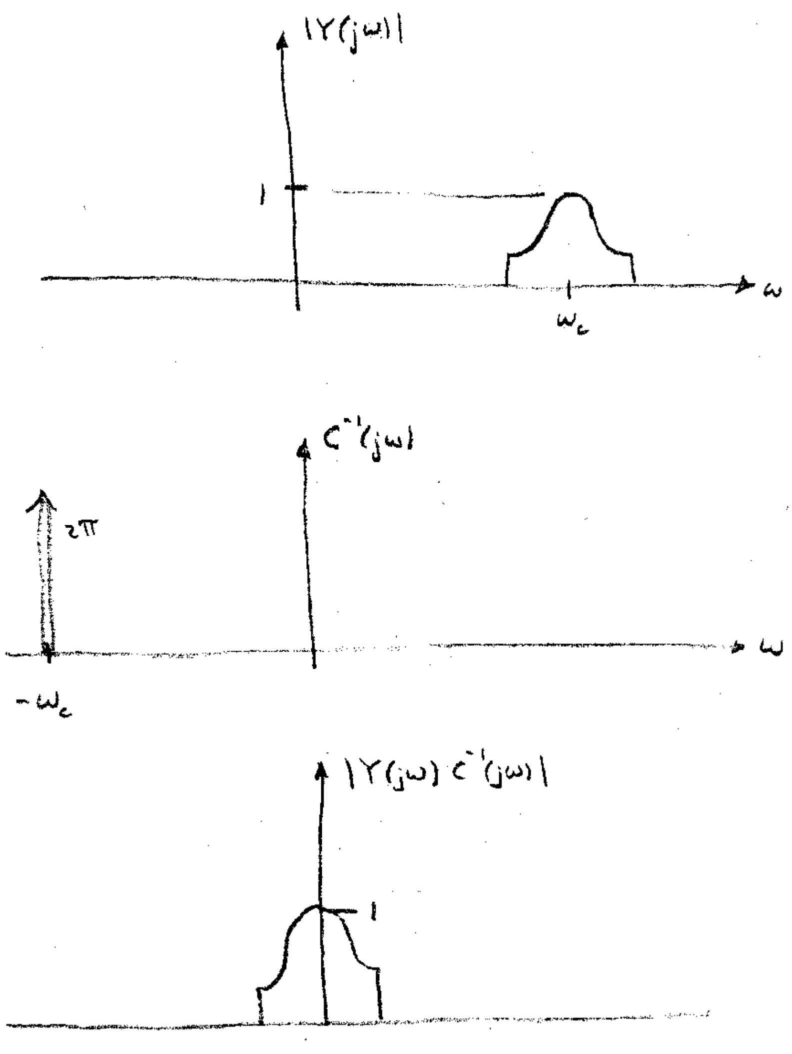


Convolving these two transforms simply results in shifting the $X(j\omega)$ transform from zero up to the carrier frequency ω_c. In particular we note that although it is shifted in frequency, the shape of $X(j\omega)$ remains unchanged.

If we want to recover the original signal *x(t)* from the transmitted signal *y(t)* we could reverse the process by multiplying by the reciprocal, or complex conjugate, of the carrier. In other words

$$x(t) = y(t)\,e^{-j\omega_c t} = y(t)\,c^{-1}(t)$$

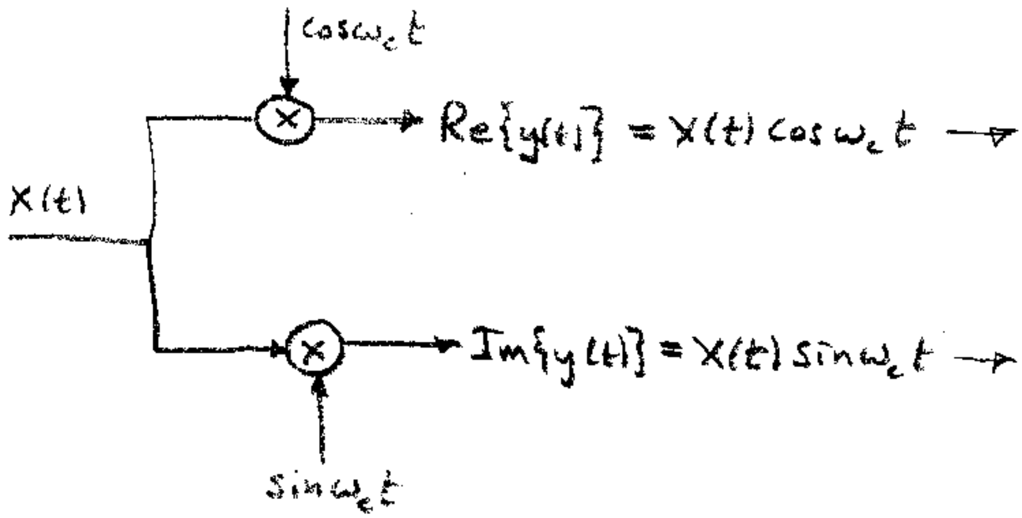
In the frequency domain this has the effect of reversing the shift of $X(j\omega)$ so that it is restored back to its original location on the frequency axis.



Actual physical implementation of this process, with hardware, requires that all of the signals must be real. Using Euler's Equation to represent the modulation of the carrier *c(t)*, we can write the modulated carrier *y(t)* as

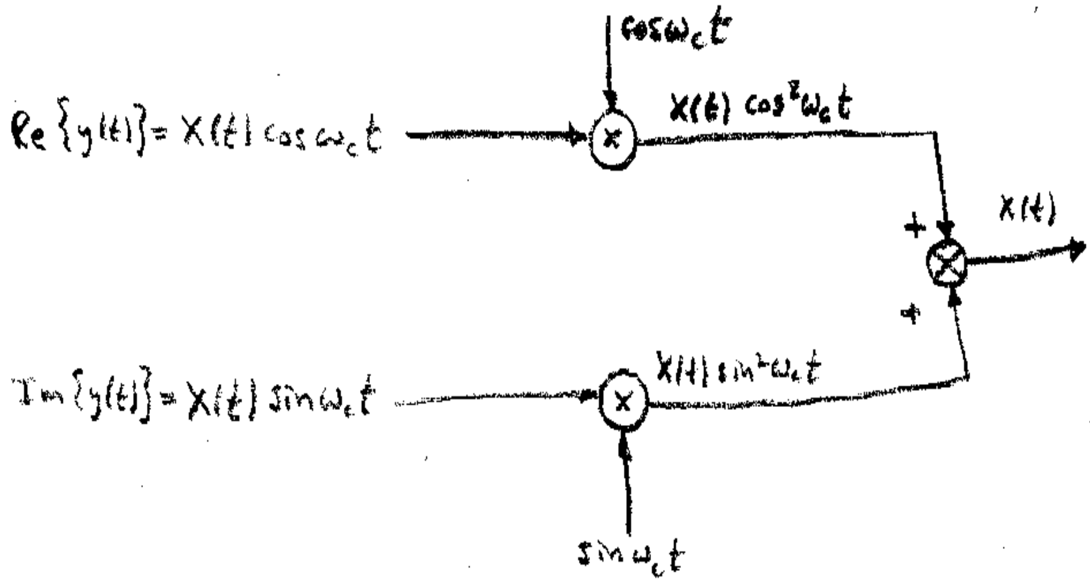
$$y(t) = x(t)\cos\omega_c t + j\,x(t)\sin\omega_c t$$

Two parallel paths must be implemented in hardware, one representing the real part of *y(t)* and the other representing the imaginary part of *y(t),* as shown in the diagram. Recall, that although one is called the real part $\Re\{y(t)\}$ and the other the imaginary part $\Im\{y(t)\}$, both of these signals are, in fact, real numbers and only *j* is the imaginary number.



These two modulated sinusoids are then transmitted through some medium (e.g., atmosphere, space, coaxial or fiber optic cable, etc.) to a receiver.

Recovering the original signal at the receiver is then accomplished by multiplying the real and imaginary parts of *y(t)* by the real and imaginary parts of the complex conjugate of *c(t)*, respectively. The results are added to recover the original *x(t)* signal.



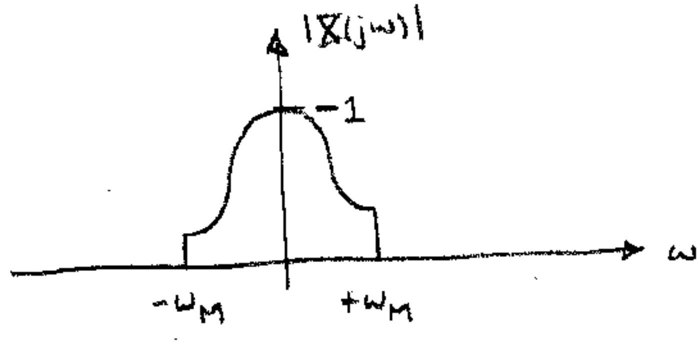
Amplitude Modulation by a Sinusoidal Carrier

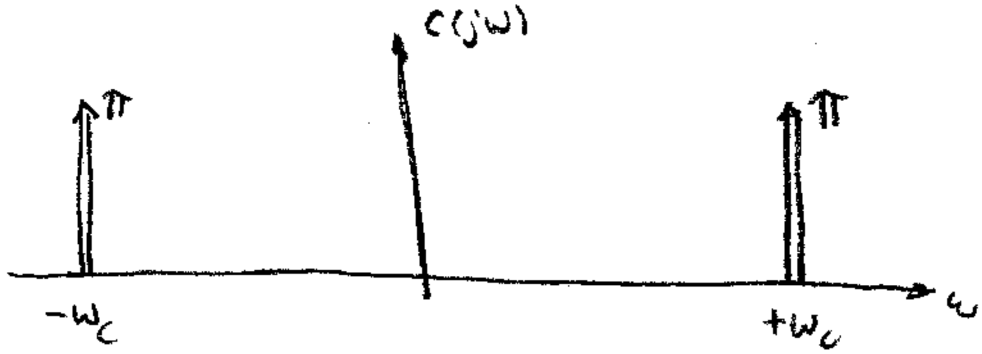
Instead of using a complex exponential carrier, which as we have seen requires transmission of two sinusoidal signals, it is possible to use only a single sinusoid. In this case we assume the following cosine form for the carrier

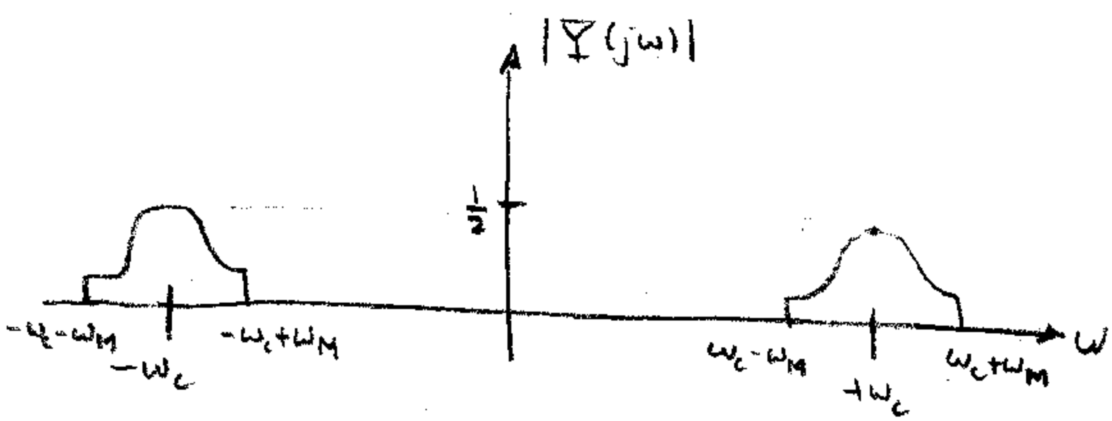
$$c(t) = \cos\omega_c t$$

and again, for convenience, we choose the time origin so that there is zero phase shift. Modulation of the carrier by the signal *x(t)* is again realized by multiplication.

Earlier we derived the Fourier transform of the cosine function as two delta functions at $\pm$ the carrier frequency. Multiplication of *x(t)* by the carrier, which again means convolution in the frequency domain, now causes the transform of *x(t)* to be replicated twice. The replicas are halved in magnitude and shifted, with one at $+\omega_c$ and the other at $-\omega_c$, as depicted in the following diagram.







This signal would then be transmitted through some medium to a receiver.

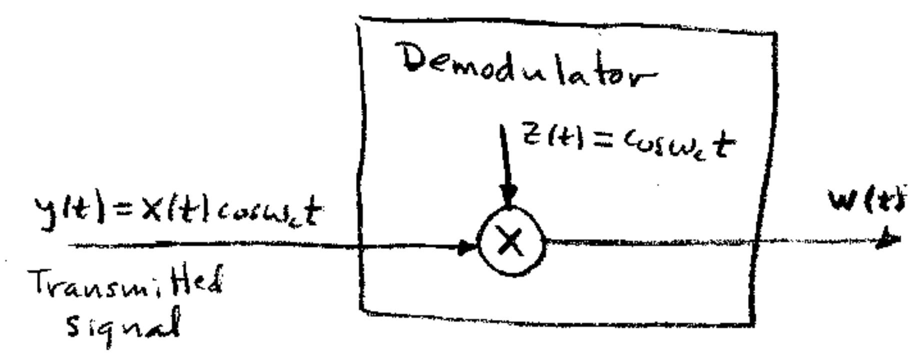
There are two common methods for demodulating the signals at the receiver to recover the original signal *x(t)*, synchronous and asynchronous demodulation. We will discuss each in turn.

Synchronous Demodulation

As the name implies, synchronous demodulation requires synchrony between the carrier of the transmitted signal (i.e., the modulated signal) and a sinusoidal signal produced by the demodulator (i.e., the demodulation signal). In particular, the transmission of the *y(t)* signal through some medium (e.g., atmosphere, coaxial cable, fiber optics, etc.) will incur some time delay. Hence, even though we postulated, at the outset, that the phase of the carrier would be zero, transmission through a medium will cause phase shift due to the time delay. Whence, there must be some synchronization mechanism at the receiver to assure that the phase of the transmitted, modulated carrier and the phase of the sinusoid produced in the demodulator are the same. Thus if *z(t)* is the sinusoid produced by the demodulator, then

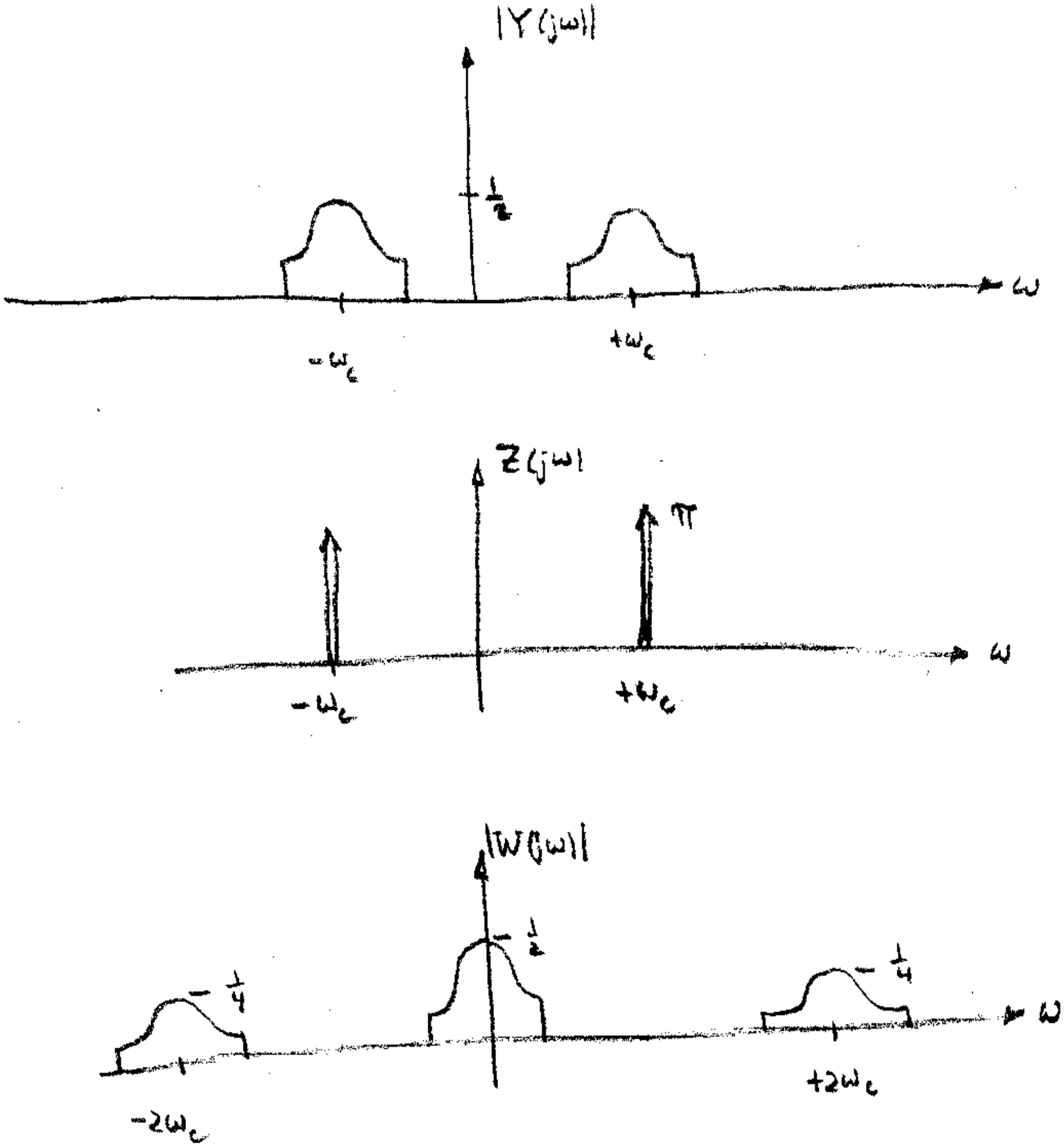
$$z(t) = \cos \omega_c t$$

where, as earlier, we now choose the time axis for the demodulator so that the phase of the synchronized carriers is zero. This signal an the received signal are then multiplied together in the demodulator to produce a signal *w(t)*



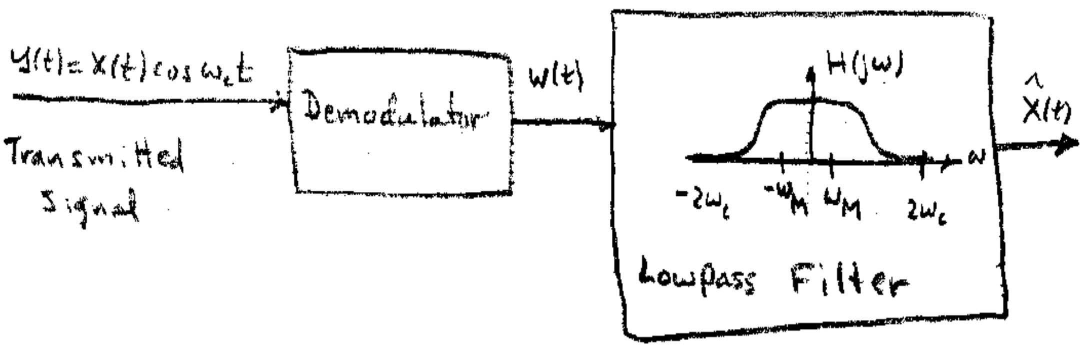
$$w(t) = x(t)\cos^2 \omega_c t$$

In the frequency domain, the effect of this multiplication of the transmitted signal *y(t)* by the sinusoid *z(t)* once again replicates the Fourier transform twice, at half the magnitude, and shifts the replicas as depicted in the following diagrams



In particular, we see that the original *x(t)* transform is now recovered in the region near the origin, but at half its original magnitude. In addition there are two replicas at $\pm 2\omega_c$, each with one fourth the original magnitude.

The original signal can then be recovered using a band pass filter centered at zero frequency, as depicted in the following diagram. The system block diagram illustrates a conceptual hardware implementation of this synchronous demodulation method



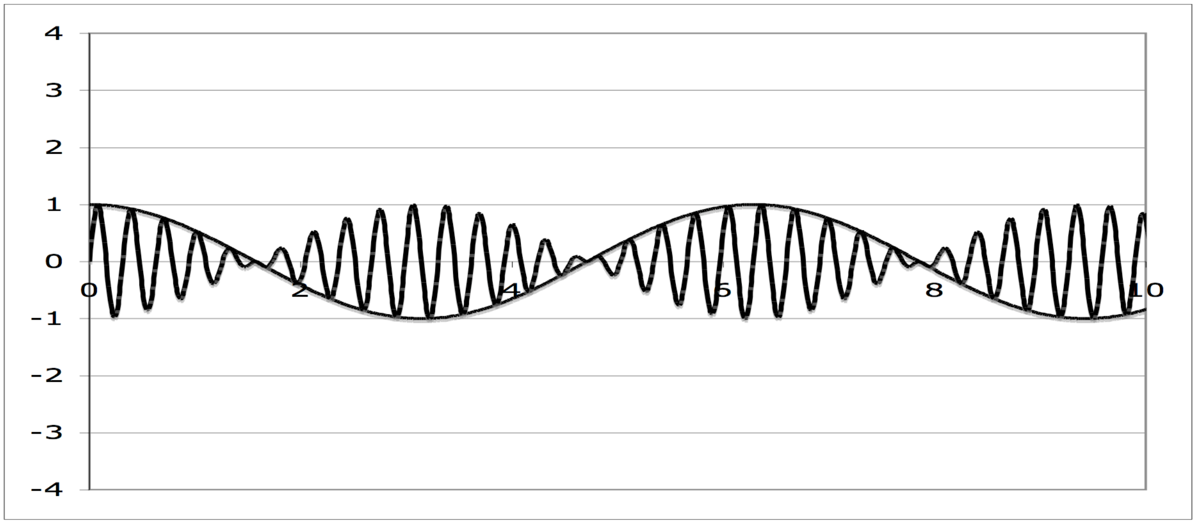
We can further illustrate this process by writing *w(t)* as

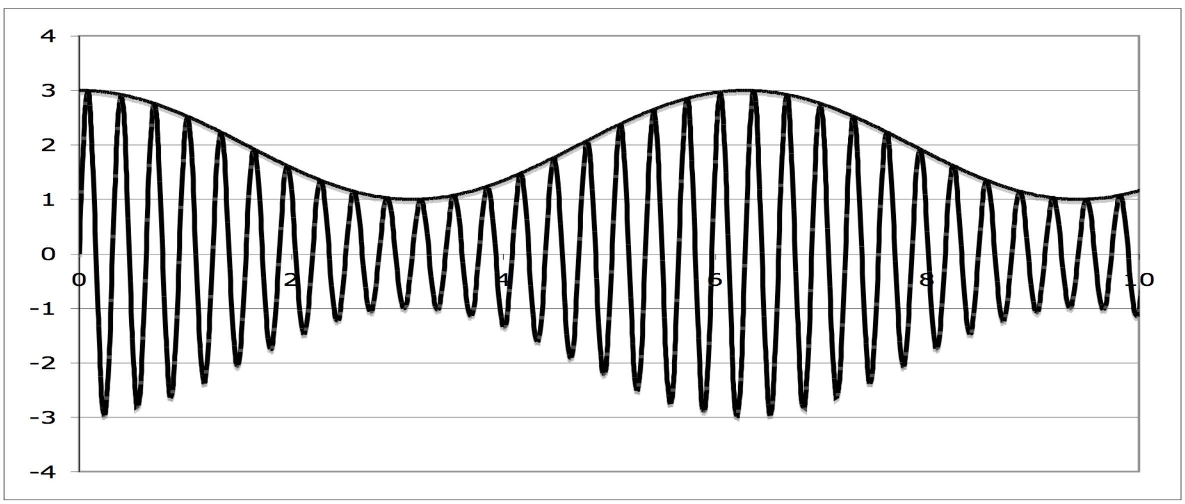
$$w(t) = x(t)\cos^2\omega_c t = \frac{x(t)}{2} + \frac{x(t)}{2}\cos 2\omega_c t$$

and realizing that the first term, with all frequency content in the rage $\pm\omega_M$, passes through the lowpass filter unattenuated, but the second term, with high frequency content, is essentially eliminated by the lowpass filter.

<u>Asynchronous Demodulation</u>
Synchronous demodulation requires some method for ensuring synchrony between the transmitted signal and the signal at the receiver that is to be demodulated. In commercial broadcasting it is important to reduce the cost of radio receivers so the simplest design is often the most desirable one. Hence, in the interests of minimizing receiver complexity, commercial amplitude modulated radio signals use asynchronous demodulation, which does not require synchrony between signals.

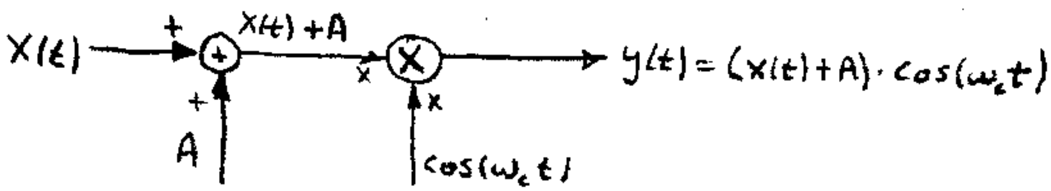
In order to understand the concept of asynchronous demodulation consider the two amplitude modulated signals illustrated below





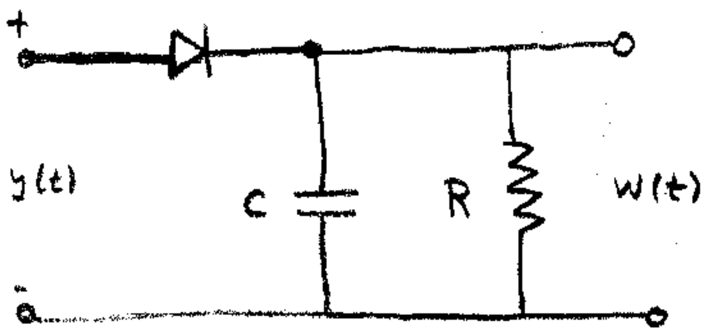
Note that in the first case the modulating signal changes sign, whereas in the second case the modulating signal is biased by a positive value *A=2*, which is large enough so that the sign of the modulating signal is always positive. In particular, note that the positive portion of the envelope of the modulated signal is a replica of the original signal. Hence, if we can detect this positive envelope we will have a replica of the original biased signal.

A modulation scheme for performing biased modulation of a signal *x(t),* to produce the modulated signal *y(t),* is illustrated in the following diagram.

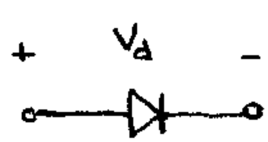


The value *A* is chosen large enough to ensure that the modulating signal (*x(t)+A)*) is always positive. The signal *y(t)* is then transmitted.

At the receiver the envelope of *y(t)* must be detected. A typical envelope detector is shown in the following diagram.



In addition to a capacitor and a resistor, the circuit also contains a diode. This nonlinear circuit element passes current in only one direction. In particular, if we depict the diode as follows-

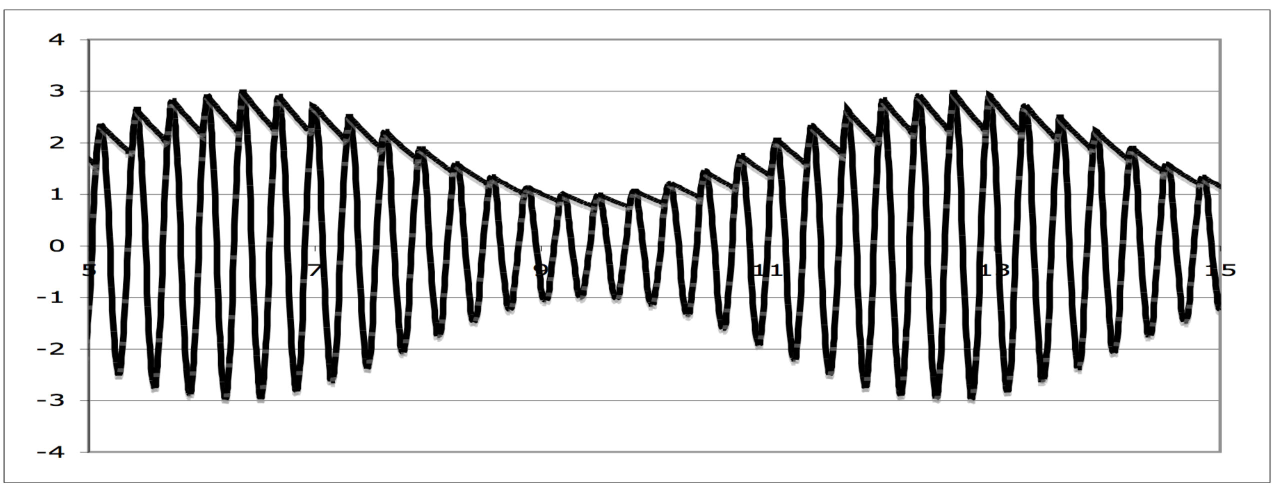


then its resistance has the following nonlinear characteristics.

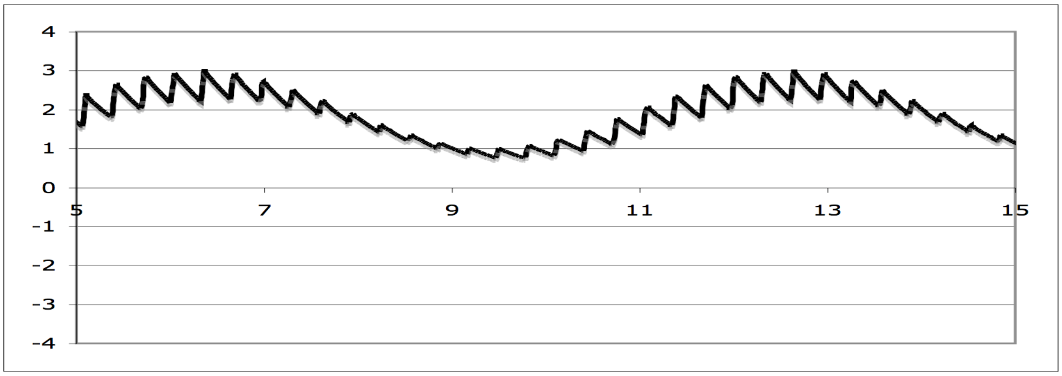
$$R_d = \begin{cases} 0 & v_d > 0 \\ \infty & v_d < 0 \end{cases}$$

At any instant of time for which the input voltage is larger than the voltage across the capacitor, then the diode will have zero resistance and the capacitor will be immediately charged so that the capacitor voltage equals the input. However, for any period of time for which the voltage cross the capacitor exceeds the input voltage, then the diode will have infinite resistance and the RC part of the circuit will be isolated from the input. In such a period the output voltage will decay as an exponential with time constant RC.

Hence, if the modulated carrier *y(t)* is input to the demodulation circuit the input and output will look like this.



And the demodulator circuit output will be



which would then be lowpass filtered to obtain a replica of the original signal, as follows.